

METHOD AND SYSTEM FOR MANAGEMENT OF STRUCTURED DOCUMENT AND
MEDIUM HAVING PROCESSING PROGRAM THEREFOR

BACKGROUND OF THE INVENTION

The present invention relates to a structured document managing method and system in which electronized document data is managed by use of a computer, and more particularly to a technique effective in the application thereof to a structured document managing system in which a structured document is managed with a logical structure of the document and an entity construction (or entity structure) thereof associated with each other.

10 The real progress of the information society has brought about the very energetic increase in amount of electronized document data generated by use of word processors, personal computers or the like. There are desired not only the mere generation or creation of various kinds of documents by a computer for use thereof for display or printout but also the realization of the great improvement in working efficiency by electronically generating documents which play an important part in the work of an organization (or mission critical documents) and
15 electronizing the whole of a work which handles these documents.

 Such a mission critical document may be made the object of various processings which include not only the mere display/printout but also the extraction of data from
25 the document, the generation of a derived document, the

reuse with a changed display style, and so forth. In many cases, therefore, the mission critical document is generated, by use of a document markup language such as SGML (Standard Generalized Markup Language), XML

5 (eXtensible Markup Language), HTML (Hyper Text Markup Language) or the like, as structured document data which are capable of being easily subjected to mechanical processing. Accordingly, a structured document managing system handling those mission critical documents has need
10 to have a function of storing and managing a structured document to perform the reference to structure information, the edition/revision of the content of the document, the management of a version history.

The structured document described by the above-
15 mentioned document markup language has a double structure which includes an entity structure and a logical structure. The logical structure is a structure which includes the order of arrangement of individual logical elements (chapters, sections, paragraphs and others) forming a
20 document and a mutual inclusion relationship between the logical elements. According to the document markup language, the logical structure of a document is represented in such a manner that a "tag" as a character string indicating the type of each logical element (or an
25 element type) is arranged at each of the head and tail of that element. The entity structure is a structure represented in such a manner that document data is divided into units called "entities" and the reference to another

entity is described in the content text. Namely, the entity structure includes a call relationship between entities.

In the case where a document becomes so long that
5 it is inadequate to maintain the document as a single continuous text, there is generally performed the management which includes, for example, dividing the document into a plurality of entities so that the respective entities are assigned to persons in charge of writing/
10 edition. Also, data such as graphic data is handled as a separate entity since it is described in a data format peculiar to each of various processing programs and is therefore incapable of co-existence with a text composed of content character strings and tags. In the case where a
15 document is formed by a plurality of entities, it is general that each entity is stored in one file. Tags used for identifying elements in a structured document, entity reference describing the reference to an entity, and character strings describing comments or the like are
20 generically termed marks.

Fig. 15 is a diagram showing an example of the conventional structured document described by a document markup language SGML. The entity structure of a single SGML document is composed of one entity corresponding to a
25 body of the document (or a document entity) and zero or one or more external entities to which the reference is made directly or indirectly from the document entity.

The external entity includes a text entity

described as an SGML text in a manner similar to the document entity, and a non-SGML data entity such as graphic data. (In the present specification, the document entity and the external entity are merely called entity in a
5 general term.) In the case of the document shown in Fig. 15 by way of example, a document entity 101 has the reference to text entities 102 and 103 and a non-SGML data entity 104 therefrom. Also, the text entity 102 includes the reference to the text entity 103 therefrom. As a
10 result, the text entity 103 having the reference from both the document entity 101 and the text entity 102 is shared between the document entity 101 and the text entity 102.

Fig. 16 is a diagram showing an example of a logical structure obtained through the analysis of the
15 conventional SGML document shown in Fig. 15. As shown in Fig. 16, a logical structure possessed by the SGML document can be represented as a tree structure which has elements and data contents as nodes.

In Fig. 16, an elliptic node represents an
20 element and a rectangular node represents data content (parsed character string data or non-SGML data). Also, an area enclosed by dotted line in Fig. 16 represents a set of nodes. A node set 201 corresponds to the document entity 101, a node set 202 corresponds to the text entity 102,
25 node sets 203 and 204 correspond to the text entity 103, and a node set 205 corresponds to the data entity 104. As shown by the relationship between the entities and the node sets, the content of an entity called by the entity

reference or an entity to be referred to is developed at a called position and a developed text or a text obtained as the result of development is subjected to syntax analysis, thereby obtaining the logical structure of the SGML
5 document.

In a typical example of the conventional technique, document data is stored in files and is managed on a file system. According to an invention disclosed by JP-A-9-223054 (hereinafter referred to as prior art 1), the
10 reading of files from storage means, the storage of files into the storage means and a version management are realized in units of a file set including a plurality of files. Though the invention of the prior art 1 primarily aims at the management of file groups forming computer
15 programs, the disclosed file management function can also be used for the purpose of managing file groups forming a large scale document. However, the invention of the prior art 1 handles each file as data having no structure. Therefore, in the case where the invention is applied to
20 the document management, it is not possible to perform an operation with the consciousness of the structure of a document.

On the other hand, an invention disclosed by, for example, JP-A-8-44718 (hereinafter referred to as prior art
25 2) is known as a prior art with which a structured document can directly be managed. A document processing apparatus disclosed by the prior art 2 includes means for analyzing the structure of a registered document to generate tree

structure data with logical elements such as chapters and sections taken as nodes and storing/managing the tree structure data. The apparatus further includes means with which a sub-tree forming a part of the document structure
5 is shared between a plurality of documents, and means with which in the case where the addition, deletion and/or updating of a logical element are made for a sub-tree, the resulting version is managed.

SUMMARY OF THE INVENTION

10 The technique shown by the prior art 1 makes it possible to manage the state of each file forming a document and the state of the document as a set of files. However, since there is no means for recognizing a logical structure possessed by the document and performing an
15 operation along the recognized logical structure, the technique of the prior art 1 is not suitable for a document managing system which handles a structured document. On the other hand, in the technique shown by the prior art 2, since tree structure data representing a logical structure
20 possessed by a registered document is directly stored and managed, the reference to structure information, the edition of structure information, the shared control in units of a sub-structure (or sub-tree), the management of versions of the sub-structure, and so forth become possible
25 for the stored data of the structured document. However, in the case where a structured document described by a document markup language such as SGML, XML or the like is

to be managed, the technique shown by the prior art 2 has the following problems to be solved.

Namely, though the apparatus disclosed by the prior art 2 is provided with the means for storing and
5 managing the logical structure of a document, there are not provided means for managing the entity structure and the logical structure in association with each other and means for performing occasional mutual conversion between relationship data which is represented in a tree structure
10 form and entity data which is described in a text form by use of a document markup language. In the case where an SGML document is to be registered in a document managing system to which the technique disclosed by the prior art 2 is applied, information concerning an entity structure
15 possessed by the SGML document before registration will be lost.

Therefore, in the case where an SGML document formed by a plurality of entities is to be registered in the document managing system applied with the technique
20 disclosed by the prior art 2 to generate logical structure data, it is not possible to establish node sets corresponding to an entity structure possessed by the document made an object of registration so that each node set is applied with an attribute (such as owner's name, access right or
25 the like) possessed by the corresponding entity and taken over therefrom. In the case where the collaboration in writing or the sharing of a sub-structure is to be made with respect to a registered document, the establishment of

a node set providing the unit of allotment or sharing and the setting of an attribute for the node set are required. However, the document managing system applied with the technique of the prior art 2 has a need to meet this
5 requirement through a manual work by a user, which makes the work remarkably complicated.

Also, in the case where a registered document is to be taken out as an SGML document in a text form again from the document managing system applied with the
10 technique of the prior art 2, it is not possible to take out document data in a form divided into a plurality of files corresponding to an entity structure existed at the time of registration. It is only possible to take out the document data as a single document entity which includes
15 the whole of document content. Accordingly, in the case where there is constructed a system in which the document managing system applied with the technique of the prior art 2 is used as a document server and the display/reuse of the document content is performed by a client computer
20 connected to the document server through a network, a partial text required on the client side is incapable of the extraction and processing thereof in units of an entity. This will result in taking out the whole document data every time. Therefore, the amount of data transferred
25 between the client and the server is increased, thereby deteriorating the efficiency of processing.

Also, in the case where there is employed a system construction in which the edition/updating of

document content is performed on the client side, it is not possible to partially update the logical structure of a registered SGML document by taking out only a specified entity portion from the document to make the edition/

5 updating thereof and registering the updated entity again.

Therefore, the co-operative management and allotted writing of a large scale document by a plurality of persons in charge are difficult.

Further, in the case where a document is to be
10 managed in such a manner that each entity forming the document is applied with digital signature data in order to ensure its content and generator (or creator), there is a problem that such strict document management is impossible in the document managing system applied with the technique
15 of the prior art 2 since text data in units of an entity is lost through the development thereof into logical structure data at the stage of registration of the document.

An object of the present invention made for solving the above problems is to provide a technique with
20 which a logical structure and an entity structure possessed by a registered document can be managed in association with each other in a mutually convertible form.

In a structured document managing method and system of the present invention for managing a structured
25 document described using a document markup language, a structured document including a plurality of entities is managed by use of relationship data which indicates an entity structure and a logical structure of the structured

document.

In the present invention, when a structured document described by a document markup language is to be registered, the document made the object of registration is
5 subjected to syntax analysis to generate syntax analysis result data having an entity structure table which indicates an entity structure of the document and parsed instance data which indicates a logical structure of the document. Referring to the generated syntax analysis
10 result data, the resource name of a resource for each entity forming the registration object document is set into the entity structure table in association with an entity identifier of that entity. Namely, entity types, entity names, entity identifiers and resource names for all
15 entities forming the registration object document are made into a table form or arranged into the entity structure table while events such as tags detected in the course of syntax analysis of document instances of the registration object document and represented by symbols or marks are
20 arranged as the parsed instance data in accordance with the order of occurrence thereof and in the form of a table indicating the correspondence between event types and parameters.

Next, the reference to the generated syntax
25 analysis result data is made to extract a relationship between an entity structure and a logical structure possessed by the registration object document so that the entity identifier and a logical structure portion obtained

by the parsed instance data are associated with each other to generate relationship data indicating the entity structure and the logical structure of the registration object document.

5 The relationship data is represented by a tree structure in which elements, data content and entity transition of the registration object document are connected as nodes. In this relationship data, that portion of sub-trees having a certain entity transition
10 node as a root which excludes all sub-trees having a lower entity transition node as a root, forms a portion included as a content in an entity corresponding to the certain entity transition node. Namely, the relationship data in the present invention has therein a relationship between
15 the entity structure and the logical structure in the form of an entity transition node.

 In the present invention, the registration of a document made an object of registration is made in a form in which the entity structures of a plurality of entities
20 originally possessed by the registration object document and logical structures thereof are held, as mentioned above. Therefore, in the case where a registered document is to be updated, it is possible to perform a processing for partial updating with a specified entity or a specified
25 logical structure portion of the registered document being selected.

 In the case where an entity of a registered document is to be updated, any entity forming the

registered document is selected as an object of updating so that the data content of the selected update object entity is subjected to updating. And, there is generated partial relationship data which indicates an entity structure and a logical structure of the update object entity. Next, relationship data of the registered document is updated by determining that portion or sub-region of the relationship data of the registered document which corresponds to the update object entity and replacing the relationship data of the determined sub-region by the generated partial relationship data of the update object entity.

In the case where a logical structure of a registered document is to be updated, any logical structure of the registered document is selected as an object of updating so that the data content of the selected update object logical structure is subjected to edition or updating. And, there is generated edition result entity data in which partial relationship data of an entity in a region of the logical structure subjected to edition is described in a character string form. The data content of the entity in the region of the logical structure subjected to edition is replaced by the generated edition result entity data. Next, a sub-region of the relationship data of the registered document corresponding to the above-mentioned entity is determined and the relationship data of the determined sub-region is replaced by the partial relationship data of that entity, thereby updating the relationship data of the registered document.

In the present invention as mentioned in the above, the entity structure and the logical structure are managed in association with each other and mutual conversion is occasionally made between relationship data
5 represented by a tree structure and entity data described in a text form by use of a document markup language. In the case where an SGML document is registered into a document managing system to which the prior art mentioned above is applied, a plurality of entities including files
10 are handled as a single document entity in which the whole of the content of the document is included. Therefore, information concerning an entity structure possessed by the SGML document before registration thereof is lost. In the present invention, however, information concerning an
15 entity structure of a document made an object of registration is held.

In the present invention, therefore, in the case where an SGML document formed by a plurality of entities is to be registered so that relationship data is generated, it
20 is possible to establish node sets corresponding to an entity structure possessed by the document made an object of registration so that each node set is applied with an attribute such as owner's name or access right possessed by the corresponding entity and taken over therefrom. In the
25 case where an allotted (or partial-charge) writing or the sharing of a sub-structure is to be made with respect to a registered document, the establishment of a node set providing the unit of allotment or sharing and the setting

of an attribute for the node set are required. In the present invention, it is not necessary for a user to make the attribute setting through a manual work since the attribute possessed by the corresponding entity is taken
5 over.

Also, in the case where a registered document is to be taken out as an SGML document in a text form again, the above-mentioned prior art is not capable of taking out document data in a form divided into a plurality of files
10 corresponding to an entity structure existed at the time of registration. In the present invention, on the other hand, the entity structure existed at the time of registration is held as it is. Therefore, it is possible to take out the document data in a form divided into a plurality of files
15 corresponding to the entity.

Accordingly, in the case where there is constructed a system in which a document managing system according to the present invention is used as a document server and the display/reuse of the document content is
20 performed by a client computer connected to the document server through a network, a partial text required on the client side can be taken out and processed in units of an entity. Namely, there will not result in that the whole document data is taken out every time. Therefore, the
25 amount of data transferred between the client and the server is reduced, thereby improving the efficiency of processing.

Also, in the case where there is employed a

system construction in which the edition/updating of document content is performed on the client side, it is possible to partially update the logical structure of a registered SGML document by taking out only a specified
5 entity portion from the document to make the edition/ updating thereof and registering the updated entity again. Therefore, the co-operative management and allotted writing of a large scale document by a plurality of persons in charge are easily possible.

10 Further, since text data in units of an entity is not lost, it is possible to make strict document management with digital signature data applied to each entity forming the document in order that its content and generator (or creator) are ensured.

15 According to the structured document managing method and system of the present invention as mentioned above, relationship data indicating each of a plurality of entities forming a document made an object of registration and a logical structure portion corresponding to that
20 entity is generated to manage the plurality of entities. Therefore, a logical structure and an entity structure possessed by a registered document can be managed in association with each other in a mutually convertible form.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Fig. 1 is a diagram showing the schematic construction of a document managing server according to an embodiment of the present invention;

Fig. 2 is a diagram showing the overall construction of a document managing system in the embodiment;

Fig. 3 is a PAD showing an epitomized procedure of a document space management program for controlling the
5 operation of the whole of the document managing server 301 in the embodiment;

Fig. 4 is a PAD showing an epitomized procedure of a document registration control program in the embodiment;

10 Fig. 5 is a diagram showing the data structure of syntax analysis result data in the embodiment;

Fig. 6 is a diagram showing a specific example of an entity structure table in the embodiment;

Fig. 7 is a diagram showing a specific example of
15 parsed instance data in the embodiment;

Fig. 8 is a diagram showing the data structure of relationship data in the embodiment;

Fig. 9 is a PAD showing an epitomized procedure of a document update control program in the embodiment;

20 Fig. 10 is a PAD showing a procedure for updating of the content of a selected entity in the embodiment;

Fig. 11 is a diagram showing an example of progress of the processing for updating of the content of the selected entity in the embodiment;

25 Fig. 12 is a PAD showing a procedure for updating of the logical structure of a selected document in the embodiment;

Fig. 13 is a diagram showing an example of

progress of a processing for updating of the relationship data of a selected document in the embodiment;

Fig. 14 is a diagram showing another example of the data structure of relationship data generated by the document registration control program in the embodiment;

Fig. 15 is a diagram showing an example of the conventional structured document described by a document markup language SGML; and

Fig. 16 is a diagram showing an example of a logical structure obtained by analyzing the conventional SGML document shown in Fig. 15.

DESCRIPTION OF THE EMBODIMENTS

In the following, the description will be made of an embodiment of a structured document managing system for managing the entity structure and the logical structure of a structured document in association with each other.

Fig. 1 is a diagram showing the schematic construction of a document managing server according to the present embodiment. As shown in Fig. 1, the document managing server in the present embodiment has a document syntax analysis processing unit 421, a data content extraction processing unit 422, a relationship data extraction processing unit 423, a data content update processing unit 424, a partial relationship data generation processing unit 425, a corresponding sub-region determination processing unit 426, a logical structure edition processing unit 427, an edition result data generation processing unit 428, a

corresponding data content update processing unit 429, an update object detection processing unit 430, and a relationship data update processing unit 431.

The document syntax analysis processing unit 421
5 is a processing unit for making the syntax analysis of a document made an object of registration, thereby generating syntax analysis result data having an entity structure table which indicates an entity structure of the registration object document and parsed instance data which
10 indicates a logical structure of the registration object document.

The data content extraction processing unit 422
is a processing unit for referring to the generated syntax analysis result data to read the data content of each
15 entity forming the registration object document so that the read data content is stored into a document space database storage area 418 and for setting the resource name of a resource of the stored entity data content into the entity structure table in association with an entity identifier of
20 the corresponding entity.

The relationship data extraction processing unit 423 is a processing unit for generating relationship data which indicates the entity structure and the logical structure of the registration object document. In the
25 relationship data, each entity forming the registration object document and a logical structure portion obtained by the parsed instance data and corresponding to that entity are associated with each other.

The data content update processing unit 424 is a processing unit with which any entity forming a registered document is selected as an object of updating from relationship data indicating an entity structure and a
5 logical structure of the registered document so that the data content of the selected entity is updated.

The partial relationship data generation processing unit 425 is a processing unit for generating partial relationship data which indicates an entity structure and a
10 logical structure of the update object entity. The corresponding sub-region determination processing unit 426 is a processing unit for determining that portion or sub-region of the relationship data of the registered document which corresponds to the update object entity.

15 The logical structure edition processing unit 427 is a processing unit for selecting any logical structure of the registered document as an object of updating from the relationship data indicating the entity structure and the logical structure of the registered document so that the
20 selected logical structure is subjected to edition. The edition result data generation processing unit 428 is a processing unit for generating edition result entity data in which partial relationship data indicating an entity existing in the region of the edited logical structure is
25 described in a character string form.

The corresponding data content update processing unit 429 is a processing unit for replacing the data content of the above-mentioned entity by the generated

edition result entity data. The update object detection processing unit 430 is a processing unit for determining that portion or sub-region of the relationship data of the registered document which corresponds to the above-

5 mentioned entity.

The relationship data update processing unit 431 is a processing unit for replacing the relationship data of the above-mentioned sub-region by the partial relationship data of the above-mentioned entity to update the relation-
10 ship data of the registered document.

It is assumed that a program for causing the document managing server to function as the document syntax analysis processing unit 421, the data content extraction processing unit 422, the relationship data extraction
15 processing unit 423, the data content update processing unit 424, the partial relationship data generation processing unit 425, the corresponding sub-region determination processing unit 426, the logical structure edition processing unit 427, the edition result data generation processing
20 unit 428, the corresponding data content update processing unit 429, the update object detection processing unit 430 and the relationship data update processing unit 431 is recorded in a recording medium such as a CD-ROM and stored in a magnetic disk or the like and is thereafter loaded
25 into a memory for execution. The medium for recording the program may be a medium other than the CD-ROM.

A document registration control program 410 performs a processing for controlling the whole of a

document registration processing, a processing for establishing node sets corresponding to an entity structure possessed by a document made an object of registration so that each node set is applied with an attribute possessed
5 by the corresponding entity and taken over therefrom, and a processing with which a digital signature applied to each entity forming the registration object document is registered together with data content of that entity.

A document update control program 412 performs a
10 processing with which in accordance with an attribute (such as owner's name or access right) of a node set established corresponding to an entity structure possessed by a registered document, the control of access to an object of updating is made for each node set to take out an object of
15 edition in units of an entity forming the registered document so that the object is subjected to updating, and a processing for controlling a processing for taking out the plurality of objects of edition separately to perform the edition of a structured document for the plurality of
20 objects of edition in parallel with each other and applying a digital signature to an entity which forms the object of edition after updating.

Fig. 2 is a diagram showing the overall construction of a document managing system in the present embodiment. As shown in Fig. 2, the document managing system in
25 the present embodiment is composed of a document managing server 310, a document managing client 302, a document work processing system 303 and a network 304.

The document managing server 301 performs a principal processing related to document management which includes the registration/storage of a document made an object of management, the reference to and the updating of
5 the structure and data content of a registered document, the search for a document having a specified attribute or character string, the deletion of a registered document, and so forth.

The document managing client 302 provides a GUI
10 (Graphical User Interface) for operating a group of documents managed by the document managing server 301. The document managing client 302 transfers a command or data for a processing such as the registration of a document, the reference to a document, or the updating or search of a
15 document to the document managing server 301 through the network 304 in accordance with an instruction from a user to cause the document managing server 301 to execute the command.

Similarly to the document managing client 302,
20 the work processing system 303 transfers a command or data to the document managing server 301 in an instruction from a user to cause the document managing server 301 to execute the command. However, the work processing system 303 provides not a general GUI for document management but a
25 dedicated GUI for performing a specified work. The network 304 is a LAN (Local Area Network) or WAN (Wide Area Network). The network 304 provides means for connecting a plurality of computers so that they transfer commands and

data to each other.

In the example shown in Fig. 2, two computers including the document managing client 302 and the work processing system 303 are used as clients for the document managing server 301. However, the client construction may be provided with only one computer or with three or more computers. Also, the client may be only the document managing client or only the work processing system.

In the example shown in Fig. 2, the network 304 is used as means for transferring commands or data from the document managing client 302 and the work processing system 303. Alternatively, there may be employed a construction in which a portable medium such as a floppy disk, magneto-optical disk, WORM (Write Once Read Multiple) optical disk or the like is used or a construction in which the portable medium and the network 304 are used in combination. Also, there may be employed a construction in which the client is implemented on the same computer as the document managing server 301 so that no data transfer is made.

Now, the description will be made of the document managing server 301 which is a main constituent element in the present embodiment. The document managing server 301 shown in Fig. 1 includes a display unit 401, a keyboard 402, a CPU (Central Processing Unit) 403, a floppy disk drive 404, a floppy disk 405, a communication controller 406, a main memory 407, a magnetic disk device 408 and a system bus 409.

The display unit 401 is used for displaying, for

example, the situation of execution of a processing in the document managing server 301. The keyboard 402 is used for inputting, for example, a command which gives an instruction for the start/stop of the server 301. The CPU 403
5 executes various programs which form the server 301. The floppy disk drive 404 is used for performing the reading/writing of data for the floppy disk 405.

The floppy disk 405 is used for performing the transfer of data (such as a document made an object of
10 registration) between the floppy disk 405 and another system. The communication controller 406 is used for communicating with the document managing client 302 and the work processing system 303 through the network 304 to perform the exchange of commands and data.

15 The main memory 407 holds various programs for performing processings in the server 301 and temporal data. The magnetic disk device 408 is used as means for storing registered document data and various data used by the server 301 for document management. The system bus 409
20 interconnects the various components or units mentioned above.

The main memory 407 has a document registration control program 410, a document reference control program 411, a document update control program 412, a document
25 output control program 413, a document search control program 414, a document space management program 415, a database managing system 416 and an operating system 417 which are held therein. The magnetic disk device 408 has a

document space database storage area 418, a program storage area 419 and a data file storage area 420 which are ensured therein.

The document registration control program 410
5 reads a document made an object of registration and described by use of a document markup language. The document is read from the floppy disk 405, the data file storage area 420 or the network 304. The program 410 makes the syntax analysis of the read registration object
10 document so that entity data and logical structure data possessed by the document and a relationship therebetween are stored into the document space database storage area 418.

The document reference control program 411 reads
15 entity data or logical structure data of a registered document stored in the document space database storage area 418 or both the entity data and the logical structure data or information to extract a part of the read data which is in turn transferred to a requesting client.

20 The document update control program 412 performs a processing for updating the content of entity data or logical structure data of a registered document stored in the document space database storage area 418 or both the entity data and the logical structure data.

25 The document output control program 413 reads entity data of a registered document from the document space database storage area 418 so that an entity forming the registered document is outputted to the floppy disk 405

or the data file storage area 420 or transferred to a requesting client (the document managing client 302 or the work processing system 303) through the network 304.

5 The document search control program 414 makes the search of a group of registered documents stored in the document space database storage area 418 in accordance with the conditional expression of search received from a requesting client and transfers search result data to the requesting client.

10 The document space management program 415 receives a command from a requesting client through the network 304 to start the document registration control program 410, the document reference control program 411, the document update control program 412, the document
15 output control program 413 or the document search control program 414 in accordance with the type of the command so that the started program performs a processing which corresponds to the content of the command. Also, the document space management program 415 performs some
20 processings which include a processing for generating a folder object for classification/adjustment of a group of registered documents to store the generated folder object into the document space database storage area 418, and a processing for deleting a part of a registered document or
25 document group from the document space database storage area 418.

 The database managing system 416 controls all processings related to the storage, updating, deletion and

search of data for the document space database storage area 418. All programs making access to the document space database storage area 418 perform their processings through the database managing system 416. The operating system 417
5 provides fundamental functions for executing programs forming the server 301 on the computer. The functions include the start/initialization of the system, the data transfer for various peripheral units, and so forth.

The document space database storage area 418 is a
10 database for storing data concerning all documents registered in the server 301 and all relevant data required by the server 301 for document management. The program storage area 419 stores therein all programs which form the server 301. These programs are read by the operating
15 system 417 at the time of system start and are transferred to the main memory 407. The data file storage area 420 stores data files which are used by the operating system 417 or various utility programs executed on the server 301.

In the present embodiment, there is employed a
20 construction in which the floppy disk 405 is used as a portable medium for data exchange or transfer. However, various portable media such as a magneto-optical disk, WORM optical disk and so forth may be used. In the present embodiment, there is employed a construction in which
25 document data and document management data are stored in the database. However, those data may be stored as files in the data file storage area 420 without a need to use the database managing system.

Next, the fundamental operation procedure of the document managing server 301 in the present embodiment will be described by use of PAD's (Problem Analysis Diagrams).

Fig. 3 is a PAD showing an epitomized procedure of the document space management program 415 for controlling the operation of the whole of the document managing server 301 in the present embodiment. When the document space management program 415 is started by a server start command or the like inputted from the keyboard 402, the program 415 enters a loop in which a predetermined processing is performed as a server each time a command from a client is received (step 501). This loop is continued until a command instructing to stop the server is inputted from the keyboard 402.

The loop of step 501 repeats a processing for receiving a command from the client (step 502) and a processing for judging the type of the received command to make a branch for a processing which corresponds to the command type (step 503).

In step 503, the type of the received command is judged to select and carry out one of steps 504 to 511 in accordance with the command type. In the case where the received command is a command making a request for registration of a new document, the document registration control program 410 is executed to store the designated document into the document space database storage area 418 (step 504).

In the case where the received command is a

command making a request for reference to data content possessed by a specified document having already been registered, the document reference control program 411 is executed to read required data from the document space database storage area 418 and to transfer the read data to the requesting client (step 505).

In the case where the received command is a command making a request for updating of data content possessed by a specified document having already been registered, the document update control program 412 is executed to update the data content of the registered document stored in the document space database storage area 418 (step 506).

In the case where the received command is a command making a request for output of a part or the whole of a specified document having already been registered, the document output control program 413 is executed to read data the content of the registered document from the document space database storage area 418 so that the read data is outputted to the floppy disk 405 or the data file storage area 420 or transferred to the requesting client through the network 304 in accordance with an instruction included in the command (step 507).

In the case where the received command is a command making a request for a set of documents satisfying a specified conditional expression of search, the document search control program 414 is executed to search a group of registered documents stored in the document space database

storage area 418 so that search result data is transferred to the requesting client (step 508).

In the case where the received command is a command making a request for generation of a folder object
5 for classification/adjustment of a group of documents, the folder object is newly generated and stored into the document space database storage area 418 (step 509).

In the case where the received command is a command making a request for association of a specified
10 document with a specified folder object or a request for deletion of a specified association, the requested addition or deletion of association is made (step 510).

In the case where the received command is a command making a request for deletion of a specified
15 document having already been registered or a specified folder object, the designated document or folder object is deleted from the document space database storage area 418 (step 511).

Fig. 4 is a PAD showing an epitomized procedure
20 of the document registration control program 410 in the present embodiment. This program 410 is called by the document space management program 415 in accordance with a command from a client making a request for registration of a document. A procedure for the operation of the document
25 registration control program 410 in the present embodiment will now be described.

When the document registration control program 410 is called by the document space management program 415,

the program 410 first calls a list of entities forming a document made an object of registration to acquire the list of entities as parameters (step 601). This entity list includes the resource name of a resource in which the data
5 content of each entity forming the registration object document is held. In the case where the registration object document is formed by a plurality of entities, the entity list further includes information for identifying which one of those entities is a document entity.

10 The resource, which is the origin of acquired data content, may include a data file stored in the floppy disk 405 or the data file storage area 420, a network resource capable of being accessed through the network 304, and a registered entity stored in the document space data-
15 base storage area 418. In the case where the resources are a data file, a resource on the network and a registered entity, its path name, its network address and an entity identifier for uniquely identifying that entity in distinction from other registered entities are designated as the
20 resource names, respectively.

Next, the document made the object of registration is assigned with a document identifier for uniquely identifying that document in distinction from all registered documents (step 602). The document syntax
25 analysis processing unit 421 makes the syntax analysis of the registration object document starting from its document entity to generate syntax analysis result data (step 603). In this syntax analysis process, those ones of entities

forming the registration object document which exclude entities having already been registered, are all assigned with entity identifiers for uniquely identifying them in distinction from all of the registered entities.

5 Next or in step 604, the data content extraction processing unit 422 refers to the generated syntax analysis result data to repeat steps 605 and 606 for all those ones of the entities forming the registration object document which exclude the registered entities. In step 605, the
10 data content of an entity made an object of processing is read from the corresponding resource. In step 606, a set of an entity identifier of that entity and the read data content thereof is stored into the document space database storage area 418 to perform entity registration.

15 Next or in step 607, the relationship data extraction processing unit 423 refers to the generated syntax analysis result data to extract a relationship between an entity structure and a logical structure possessed by the registration object document, thereby
20 generating relationship data which indicates each entity forming the registration object document and a logical structure portion corresponding to that entity. In step 608, a set of the document identifier assigned to the registration object document and the generated relationship
25 data is stored into the document space database storage area 418 to perform document registration, thereby completing the process.

Fig. 5 is a diagram showing the data structure of

the syntax analysis result data in the present embodiment. Referring to Fig. 5, the syntax analysis result data 701 generated by the document registration control program 410 in step 603 includes two principle parts, that is, an
5 entity structure table 702 and parsed instance data 703.

The entity structure table 702 includes data in which the entity type, entity name, entity identifier and resource name of each entity forming a document made an object of registration are arranged in a table form. The
10 item "ENTITY TYPE" uses a symbol to indicate whether the corresponding entity is a document entity (DOC), a text entity (TEXT) or a non-SGML data entity (NDATA). The item "ENTITY NAME" uses a character string to indicate an entity name which is used in the registration object document for
15 referring to the corresponding entity. However, in the case where the entity is a document entity, this item is blank. The item "ENTITY IDENTIFIER" indicates an entity identifier assigned to the corresponding entity. The item "RESOURCE NAME" uses a character string to indicate the
20 resource name of a resource holding the data content of the corresponding entity. However, in the case where the entity is a registered entity, this item is blank.

The parsed instance data 703 is data in which events detected in the course of syntax analysis of
25 document instances of the registration object document are arranged in accordance with the order of occurrence thereof and in a table form indicating the correspondence between event types and parameters. The item "EVENT TYPE" uses a

symbol to indicate whether the corresponding event is a start tag (STAG) of an element, an end tag (ETAG) thereof, a tag (ITAG) serving as both the start and the end of an element having no content, parsed character string data (PCDATA), entity reference (EREF) or entity end (EE). The item "PARAMETER" indicates additional data determined corresponding to the event type. In the case where the event type is a start tag, the element type name of an element having that start tag at a head position and an attribute value (option) designated in the start tag are indicated as a parameter. In the case where the event type is an end tag, the element type name of an element having that end tag at a tail position is indicated as a parameter. In the case where the event type is parsed character string data, the content of the character string is indicated as a parameter. In the case where the event type is entity reference, the entity name of an entity to be referred to is indicated as a parameter. The entity end event has no parameter.

Fig. 6 is a diagram showing a specific example of the entity structure table 702 in the present embodiment. The example shown in Fig. 6 corresponds to an entity structure table 702 generated as the result of syntax analysis of the SGML document shown in Fig. 15. Rows 801 to 804 in Fig. 6 respectively correspond to the document entity 101, the text entities 102 and 103 and the data entity 104 shown in Fig. 15.

Fig. 7 is a diagram showing a specific example of

the parsed instance data 703 in the present embodiment.

The example shown in Fig. 7 corresponds to parsed instance data 703 generated as the result of syntax analysis of the

SGML document shown in Fig. 15. As shown in Fig. 7, the

5 SGML document described in accordance with the syntax of

the document markup language SGML is converted into a

sequence of events which indicates the occurrence of tags,

parsed character string data, entity reference and so forth

(step 603 of Fig. 4 mentioned above).

10 The relationship data extraction processing unit

423 of the document registration control program 410 refers

to the parsed instance data 703 of the syntax analysis

result data to extract a logical structure possessed by the

registration object document and associates entity refer-

15 ence (EREF) in the parsed instance data 703 with an entity

identifier in the entity structure table 702, thereby

generating relationship data in which the entity identifier

of each entity forming the registration object document and

a logical structure portion corresponding to that entity

20 are represented by a tree structure.

Fig. 8 is a diagram showing the data structure of relationship data in the present embodiment. Fig. 8 shows,

by way of example, the data structure of relationship data generated by the relationship data extraction processing

25 unit 423 of the document registration control program 410

in the case where the SGML document shown in Fig. 15 is

processed as a document made an object of registration.

As shown in Fig. 8, the relationship data 1001 is

represented as a tree structure. In Fig. 8, an elliptic node represents an element and a character string in this node indicates the element type of that element. A rectangular node represents data content and a character string in this node indicates the type of the data content (parsed character string data PCDATA or non-SGML data NDATA). A rhombic node represents an entity transition and a character string in this node indicates an entity identifier.

10 That portion of sub-trees having a certain entity transition node as a root which excludes all sub-trees having lower entity transition nodes as roots, is included as a content in an entity corresponding to that certain entity transition node. Namely, the relationship data
15 generated by the document registration control program 410 in the present embodiment includes a relationship between a logical structure and an entity structure (or entity construction) in the form of an entity transition node. In the example shown in Fig. 8, a sub-region 1002 enclosed by
20 chained line corresponds to the content of the entity E0002 (or the text entity 802 shown in Fig. 6).

Fig. 9 is a PAD showing an epitomized procedure of the document update control program 412 in the present embodiment. This program 412 is called by the document
25 space management program 415 in accordance with a command from a client making a request for updating of a document. A procedure for the operation of the document update control program 412 in the present embodiment will now be

described.

When the document update control program 412 is called by the document space management program 415, the program 412 first judges whether the content of any entity
5 forming a registered document or a logical structure of the registered document is designated as an object of updating (step 1101). In the case where the entity is designated as the object of updating, the flow branches to processings in steps 1102 and 1103. On the other hand, in the case where
10 the logical structure is designated as the object of updating, the flow branches to processings in steps 1104 and 1105. At the time of return from the branch, the process is completed.

In step 1102, a parameter (or an entity
15 identifier) given at the time of call is used to select the entity of a registered document which is made an object of updating. Next or in step 1103, a processing for updating the data content of the selected entity is performed.

In step 1104, a parameter (or a document
20 identifier) given at the time of call is used to select a registered document which holds a logical structure made an object of updating. Next in step 1105, a processing for updating the logical structure possessed by the selected document is performed.

25 Fig. 10 is a PAD showing a procedure for updating the content of a selected entity in the present embodiment. The details of step 1103 in Fig. 9 or the processing for updating the content of the selected entity will now be

described.

As shown in Fig. 10, the data content update processing unit 424 for performing the updating of the selected entity first sets a lock for the update object
5 entity (step 1201). Such lock setting is made in order to prevent one entity from being updated by a plurality of users. Next, the check is made of whether or not the lock setting results in a success (step 1202). In the case of a success, an update processing in steps 1203 to 1211 is
10 performed. In the case of a failure, an error code indicating that the lock has already been set and the updating is therefore impossible is sent back (step 1213), thereby completing the process.

In step 1203, the data content of the update
15 object entity is read from the document space database storage area 418. Next or in step 1204, the read data content is outputted to an external area (such as the floppy disk 405, the data file storage area 420, or a work area on the client side connected through the network 304)
20 in which the edition of the data content using an external program is possible. And, the process is once interrupted to wait for the completion of edition of the outputted data content by the external program (for example, an SGML editor). When a notice of completion of the edition
25 processing by the external program is received, the process is restarted so that the flow goes to step 1205. In step 1205, the data content after updating is read from the external area and the data content of the update object

entity is replaced by the read data content which is in turn stored into the document space database storage area 418.

Next or in step 1206, the partial relationship
5 data generation processing unit 425 judges the type of the update object entity. In the case where the update object entity is a document entity or a text entity, steps 1207 and 1208 are carried out. In the other case or in the case where the update object entity is a non-SGML data entity,
10 step 1209 is carried out.

In step 1207, the syntax analysis of the updated data content of the update object entity is made to generate syntax analysis result data which indicates an entity structure and parsed instance data of the update
15 object entity. In step 1208, the reference to the syntax analysis result data is made to generate partial relationship data in which each entity forming the update object entity and a logical structure portion obtained by the parsed instance data are associated with each other.

20 In step 1209, the updated data content itself is set as the partial relationship data.

Next or in step 1210, the corresponding sub-region determination processing unit 426 searches the relationship data of all registered documents to detect all
25 entity transition nodes which correspond to the update object entity (or hold the entity identifier of the update object entity). For all of the detected entity transition nodes, step 1211 is repeatedly carried out. In step 1211,

the relationship data update processing unit 431 replaces partial relationship data at a position lower than the selected entity transition node by that partial relationship data set in step 1208 or step 1209 which is in turn
5 stored into the document space database storage area, thereby updating the relationship data of the registered document corresponding to the updated entity. Finally or in step 1212, the lock set for the update object entity is released. Thereby, the process is completed.

10 Fig. 11 is a diagram showing an example of the progress of a process for updating of the content of a selected entity in the present embodiment. Fig. 11 represents a specific example of the progress of the updating of the content of the selected entity the procedure of
15 which has been shown in Fig. 10. The example shown in Fig. 11 corresponds to the case where the content of the entity text 103 shown in Fig. 15 is outputted to an external area so that it is subjected to edition by an external program.

In the shown example, it is assumed that a part
20 of a content character string included in an element "note" in the text entity 103 is enclosed by tags and so that it is changed into an independent element "em". With this edition processing, the content of the text entity 103 is updated as indicated by edition result entity
25 data 1301. Since this entity is a text entity, syntax analysis is carried out in step 1207 shown in Fig. 10 and partial relationship data is generated in step 1208. In the case of the shown example, there results in that the

generated relationship data has a structure indicated by relationship data 1302.

Next or in step 1210, a search is made for an entity transition node corresponding to the update object entity. In the example shown in Fig. 11, the registered document has the relationship data 1001 shown in Fig. 8. Therefore, two entity transition nodes (at upper extremes of shaded node in the relationship data 1001 shown in Fig. 11) corresponding to the updated text entity 103 (or the entity identifier E0003) are detected on this tree structure. Accordingly or in step 1211, the replacement of sub-structures having those entity transition nodes as roots is made so that the relationship data of the document is ultimately updated as indicated by relationship data 1303.

Fig. 12 is a PAD showing a procedure for updating the logical structure of a selected document in the present embodiment. The details of step 1105 in Fig. 9 or the processing for updating the relationship data of the selected document will now be described.

As shown in Fig. 12, the logical structure edition processing unit 427 for performing the updating of the logical structure of a selected document first selects an edition region (step 1401). The selection of the edition region is made by designating any node on the logical structure of relationship data of a document to be subjected to edition. Sub-trees having the designated node as a root form a region for which the edition is possible.

Next or in processings of steps 1402 to 1406, locks are set for all entities corresponding to the selected edition region in order to prevent the duplicate updating by a plurality of users. The entities corresponding to the selected region include entities which correspond to entity transition nodes included in the selected region and an entity transition node existing at a position upper than the selected region and nearest thereto.

10 In step 1403, an updating lock is set for a specified entity which is now under consideration. Next, the check is made of whether or not the lock setting results in a success (step 1404). In the case of a failure, all presently set locks are released and there is
15 then sent back an error code indicating that an updating lock has already been set and the updating is therefore impossible (step 1406). Thereby, the process is completed.

In the case where the setting of locks for all the corresponding entities results in a success, the flow
20 goes to step 1407. In step 1407, an editing operation for the logical structure of the relationship data (the addition/deletion of a lower node, the change of character string data, and so forth) is performed in accordance with a designation from a user through a GUI or the like. At
25 this time, an entity including a position subjected to change or edition is applied with an update mark indicating that the entity has the changed content.

When the completion of the editing operation is

indicated, the flow goes to step 1408. In step 1408, processings of steps 1409 to 1414 are repeated for all entities applied with the update marks in the stage of the editing operation.

5 In step 1409, the edition result data generation processing unit 428 acquires partial relationship data for a region corresponding to an entity which is now under consideration. The acquired partial relationship data is stored into a temporary storage. In step 1410, there is
10 generated edition result entity data in which the partial relationship data acquired in step 1409 is described in a character string form in accordance with the syntax of the document markup language SGML. Next or in step 1411, the corresponding relationship data update processing unit 429
15 replaces the data content of the entity under consideration by the generated edition result entity data which is in turn stored into the document space database storage area 418.

 In step 1412, the update object detection
20 processing unit 430 searches the relationship data of all registered documents to detect all entity transition nodes which correspond to the updated entity under consideration (or hold the entity identifier of the entity under consideration). For all of the detected entity transition
25 nodes, processings of steps 1413 and 1414 are repeatedly performed.

 In step 1413, the relationship data update processing unit 431 judges whether or not an entity

transition node under consideration is a node corresponding to a region having already been subjected to the edition of the logical structure in step 1407. In the case where the entity transition node under consideration is not such a node or has not yet been edited, a copy of the partial relationship data after updating acquired in step 1409 is generated and a sub-tree having the entity transition node under consideration as a root is replaced by the generated copy (step 1414).

10 When the repetition in step 1408 is completed, the flow goes to step 1415 in which the relationship data for all documents subjected to change is stored into the document space database storage area 418. Finally, the locks set for the entity group are all released (step 15 1416), thereby completing the process.

Fig. 13 is a diagram showing an example of the progress of a process for updating of the relationship data of a selected document in the present embodiment. Fig. 13 represents a specific example of the progress of the updating of the relationship data of the selected document the procedure of which has been shown in Fig. 12. The example shown in Fig. 13 corresponds to the case where the relationship data of the SGML document shown in Fig. 15 is edited.

25 Relationship data 1001 shown in Fig. 13 corresponds to a logical structure before the edition is made. The relationship data 1001 is the same as the relationship data 1001 shown in Fig. 8. Now provided that the selection

of an edition region is made by designating a "sec" type element 1501, sub-trees having the "sec" type element 1501 as a root form a region for which the edition is possible. As a result, an entity E0003 corresponding to an entity
5 transition node included in the edition region and an entity E0002 corresponding to an entity transition node at a position upper than the edition region and nearest thereto are subjected to the setting of locks as entities corresponding to the edition region.

10 After the lock setting, an editing operation is performed. In the shown example, the structure edition is made in such a manner that parsed character string data existing at a position lower than a "note" type element 1502 is divided so that a part thereof forms an "em" type
15 element. In this case, the logical structure of the whole document after updating results in relationship data 1503. Provided that only the above editing operation is ultimately performed, an entity applied with a mark as an entity subjected to content change is only E0003. Edition
20 result entity data is generated from partial relationship data after updating by configuring it into a character string form in accordance with the syntax of the document markup language SGML. In the shown example, the same character string data 1301 as the edition result entity
25 data 1301 shown in Fig. 11 is obtained as the edition result entity data.

 After the entity E0003 is updated by replacing the data content thereof by the edition result entity data

1301, each of structures at positions lower than all entity transition nodes corresponding to the updated entity E0003 (but excluding the entity transition node included in the region subjected to structure edition) is replaced by a
5 copy of the partial relationship data after structure edition. In the shown example, since another entity transition node corresponding to the entity E0003 is included in the document, the replacement of the structure of this portion is made. Ultimate relationship data 1303
10 after updating results in the same as the relationship data 1303 shown in Fig. 11.

As shown in the above, a document update processing in the present embodiment results in that whether an entity forming a document is edited as a text or the
15 logical structure of the document is directly subjected to structure edition, the same result having a consistency or matching between the logical structure and the data content of the entity remains ultimately if the substantial content of edition is equivalent. Though the present embodiment
20 has been described in the case where a structured document described by use of SGML is processed as a document made an object of registration, a structured document described by use of another document markup language such as XML, HTML or the like may be made the object of registration.

25 Fig. 14 is a diagram showing another data structure of relationship data generated by the document registration control program 410 in the present embodiment. Fig. 14 shows, by way of example, relationship data

generated in the case where the SGML document shown in Fig. 15 is processed as a document made an object of registration. This other example of relationship data in the present embodiment will now be described.

5 As shown in Fig. 14, relationship data 1601 does not include entity transition nodes in the tree structure, unlike the relationship data 1001 shown in Fig. 8. In the relationship data 1601, an entity transition node is not held as an independent node. Thereinstead, individual
10 element nodes and data content nodes forming the logical structure hold the entity identifiers of the corresponding entities as properties.

 Since all the element nodes and the data content nodes have the entity identifiers, the data amount of the
15 relationship data 1601 is large as compared with that of the relationship data 1001 shown in Fig. 8. However, when an entity corresponding to an edition region is to be determined in the processing for updating of the logical structure of the selected document shown in Fig. 12, a need
20 to search for entity transition nodes with the upward tracing of the tree structure of nodes is eliminated but the acquisition of entity identifiers held as properties by nodes included in the edition region suffices, thereby simplifying the processing and improving the efficiency of
25 processing. No functional difference is caused between the case where the relationship data 1601 is used and the case where the relationship data 1001 is used.

 According to the structured document managing

system in the present embodiment as described above,
relationship data indicating each entity forming a document
made an object of registration and a logical structure
portion corresponding to that entity is generated to manage
5 a plurality of entities. Therefore, the management of a
registered document can be made with a logical structure
and an entity structure of the document associated with
each other in a mutually convertible form.

According to the structured document managing
10 system in the present embodiment, an attribute originally
possessed by an entity forming a registered document is
taken over. Therefore, it is possible to make the control
of access to each entity of the registered document without
making the manual setting of attribute, that is, in
15 accordance with the attribute possessed by each entity of a
structured document before registration.

According to the structured document managing
system in the present embodiment, the processing of a
registered document is performed while partially taking out
20 the document in units of an entity. Therefore, it is
possible to reduce the amount of data transferred at the
time of entity taking-out, thereby improving the efficiency
of processing.

According to the structured document managing
25 system in the present embodiment, the edition is made while
separately taking out a plurality of objects of edition
from a registered document in units of an entity. There-
fore, it is possible to easily make the co-operative

management and allotted writing of a large scale document by a plurality of persons in charge.

According to the structured document managing system in the present embodiment, there is no fear that
5 text data in units of an entity is lost. Therefore, it is possible to make the strict management of a document with digital signature data applied for each entity forming the document in order to ensure the content and generator (or creator) of that entity.

10 According to the present invention, since relationship data indicating each entity forming a document made an object of registration and a logical structure portion corresponding to that entity is generated to manage a plurality of entities, it is possible to manage a logical
15 structure and an entity structure of a registered document in such a manner that they are associated with each other in a mutually convertible form.